DOCUMENT RESUME

ED 422 605 EA 029 140

AUTHOR	Flanders, Anne K.; Wick, John
TITLE	Individual and Organizational Learning during Systemic School Improvement.
PUB DATE	1998-04-00
NOTE	46p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998). Document contains light type.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Accreditation (Institutions); *Change Strategies; Educational Assessment; *Educational Change; *Educational Improvement; Elementary Secondary Education; Outcomes of Education; Time Factors (Learning)
IDENTIFIERS	*Systemic Educational Reform

ABSTRACT

To explore the degree to which schools change, an analysis of how systemic participants (external agencies, the school, and individuals in the school) facilitate or constrain the rate of change is offered. Change, or learning, was assessed by following schools' and reviewers' use of an accrediting agency's criteria for school improvement, in this case, expectations for the improvement of student learning set by schools initiating Outcomes Accreditation (OA) with the North Central Association. OA is based on meeting criteria for the improvement of student learning. Therefore, setting rigorous goals for improved learner outcomes was the primary focus of schools' activities leading up to implementation in the school. Schools participating in field testing prior to the formalization of OA guidelines provided a baseline for comparisons. Assimilation and use of OA criteria in goal statements and peer evaluations were assumed to be an indicator of organizational and individual learning. Three questions guided the research: Did the schools and peer reviewers learn to use the criteria? How does change in use of the criteria develop? and Did school characteristics or factors in the school's environment facilitate or constrain the rate of change? Results revealed that time was the most important ingredient for improved use of the criteria by schools and by reviewers. Two appendices feature improvement target areas and a template used to assess goals and compliance. (Contains 22 references.) (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 605

Individual and Organizational Learning During Systemic School Improvement

by
Anne K. Flanders and John Wick
Northwestern University
Evanston, Illinois

Schools often are considered by policy implementors to be resistant to the very policies, research, and theories that are assumed to improve student learning. If implementing good change is so difficult in schools, then it seems logical that studying implementation might offer direction that can facilitate more successful educational improvement in the future. Exploring the progress of intentional change, especially how it is influenced by school contexts and their systemic relationships, offers an opportunity to describe what policy initiators perceive is resistance to change in schools from another perspective.

A change in voluntary accrediting practices that required participants to change what they were doing in schools offered an opportunity to explore this area. This investigation focused on expectations for the improvement of student learning set by schools initiating Outcomes Accreditation (OA) with the North Central Association (NCA) from 1992 through 1994 and evaluations these expectations by professional peers. After a period of several years schools were able to use OA criteria for school improvement and reviewers applied the criteria in evaluative judgments. Environmental pressure and support to change, school autonomy to deal with changing, the degree criteria required a departure from conventional practice, and individual and organizational activism were significant in affecting the rate with which schools and reviewers used OA guidelines for school improvement.

EA 029 140

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

G. Flanders

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The Problem

Recently, educational reforms that include agencies and persons in the work of improving learner outcomes in schools have received support for their potential to link policy and practice. Because these processes intentionally engage school personnel with outside organizations, individuals, and agencies they are often referred to as "systemic" changes or reforms. Reports on the progress of state systemic reforms and other school improvement initiatives have begun to appear. They highlight the complexity and difficulty of developing change in schools that is intent on improving learning. However, descriptions of the progress of school improvement efforts guided by professional organizations are rare.

The study was conducted to get a better understanding of the course of change processes, and how systemic participants (external agencies, the school, and individuals in the school) facilitate or constrain the rate of change. Of particular interest was the impact of the site for changing, and unseen systemic connections that might influence the school's change outcomes. To follow the initiation of change, school improvement plans submitted for OA, and peer evaluations of these plans were tracked from 1992 through 1994.

Change, or learning, was assessed by following schools' and reviewers' use of the accrediting agency's criteria for school improvement. Outcomes Accreditation is based on meeting criteria for the improvement of student learning. Therefore, setting rigorous goals for improved learner outcomes was the primary focus of schools' activities leading up to implementation in the school. Schools seeking accreditation from NCA are required to document their improvements at the end of a three to five year cycle. What they choose to improve is left up to them (NCA 1994).

Schools participating in field testing (from 1987-1991) prior to the formalization of OA guidelines for school improvement provided a baseline for comparisons. Assimilation and use of OA criteria in goal statements and peer evaluations was assumed to be an indicator of organizational and individual learning. The influence of school and environmental characteristics on a school's use of criteria were explored, as was the influence of environmental and individual factors on reviewers' evaluations of the plans. Three questions guided our research:

- Did the schools and peer reviewers learn to use the criteria?
- How does change in use of the criteria develop?
- Did school characteristics or factors in the school's environment facilitate or constrain the rate of change?

Perspectives

Cognitive science has described human learning as a constructive process leading to the development of meaningful, useful knowledge (Moffett 1994, Bruer 1993). During the process of learning, persons incrementally build upon their current knowledge. Application of new knowledge in real situations, introspection, and reflection with others are also important. Therefore, the full development of personal knowledge involves individual and social processes in addition to the acquisition of skills and facts.

Two decades ago, DeBoard commented that organizational learning was, "individual behavior writ large." (DeBoard p.24, 1978) or similar to individual learning behavior. The parallel holds today as more is understood about learning through cognitive science and the studies of organizational behavior. In fact, Senge (1990) uses a "constructivist" three-stage model of learning to describe systemic change processes in organizations. According to this model organizations move

incrementally from the current state toward the changed state until they fully incorporate new beliefs.

The analogy for learning hold true for schools and educators changing accustomed practices. In 1987, reflecting on the Rand Change Agent Study ten years earlier, McLaughlin concluded that successful school change required organizational capacity for dealing with the stress of changes and engagement of human will. In 1994 she added the concept of learning, or "getting it" (McLaughlin 1994), as a third condition. The "learning" of change in educational organizations can be seen in descriptions of teachers attempting to change practice (Knapp, Shields, and Turnbull 1995) and schools attempting to improve student learning (Louis and Miles 1990). So prevalent is this pattern that Tyack and Cuban (1995) have concluded that policies should be evaluated and modified based on their usefulness in practice rather than practitioners willingness to use them.

Tyack and Cuban (1995) also agree that organizational learning can be advanced or constrained by systems that are interdependent with the school. In fact, the involvement of external agencies is vital to Rothman (1995) and Schlechty (1997) who believe that change in schools is fragile unless it is linked to support from the school's environment. Others disagree. Meier (1998), for one, feels external agents divert schools from carrying out their mission - to educate students. Like Tyack and Cuban she thinks that large scale, top-down efforts with systemwide effects are really very "unsystemic" in nature because they ignore the response from schools who must implement these policies.

Regardless, the basis for a systemic explanation of organizational change incorporates natural and ecological theories for group behavior. Borrowing from the natural model of group behavior, systemic change is thought to be dependent upon what the change requires, the context where it takes place, and the person(s)

involved (Ross and Nisbett, 1991). The influence of ecological forces is also added to this mix (Senge 1990).

Four beliefs about systemic change emerge from the combination of natural and ecological perspectives. First, systems are limited in their capacity to address change (McLaughlin 1987, Lewin 1951). Second, pressure for change produces an opposite reactions to regain systemic equilibrium (Ross and Nisbett 1991, Lewin 1951). Third, all parts with systemic connections need to be engaged in creating and using change (Kanter, Stein, and Jick 1992). Finally, no change is complete in theory or policy, this can only come with action (Tyack and Cuban 1995, Argyris 1993).

Because they already incorporate systemic responsiveness, professional organizations have received attention for their potential to advance large-scale, educational change (Tyack and Cuban 1995, Darling-Hammond and McLaughlin 1995). The ability they have to protect their membership from the whims of political agendas, according to Wagner (1993), provides the necessary stability for schools to focus on their own improvement. Moffett (1994) points out that these organizations also have greater potential to facilitate cooperation because their membership bridges conventional organizational boundaries. In comparison to hierarchical, top-down superordinate agencies, they are believed to function with greater flexibility due to their flattened organizational structures (Darling-Hammond and McLaughlin 1995).

One current example of systemic reform in a professional organization is a shift in voluntary accreditation practices in one of the regional voluntary accrediting agencies. The North Central Association (NCA) began field testing a new form of school accreditation in the middle 1980s called Outcomes Accreditation (OA) (Wick and Sarterfiel 1992). Outcomes Accreditation is offered to schools willing to

undertake a three to five year effort resulting in documented improvements in student learning. In 1992 the criteria for OA school improvement was formally set and used in peer evaluation by NCA. This criteria was based on almost a decade of debate over theories of how schools improve, successful practices of schools improving learner outcomes, and national and international comparisons of student achievement (Wick and Gose, 1994). By 1993 almost one fifth of the 8,000 NCA secondary and elementary member schools had selected OA for their accreditation cycle (Wick and Gose 1994).

Methods

Subjects

Schools

The subjects of this study were 637 schools electing to pursue NCA Outcomes Accreditation between 1987 and 1994. Because criteria for schools' sets of improvement goals was not formalized by NCA until 1992; plans taken from schools participating in field testing (prior to 1992) are treated a single grouping. Table 1 provides a breakdown of the number of schools submitting plans, the number of states in the NCA region represented, the average number of goals per plan, and the total number of goals submitted by schools in their plans.

None of the schools were selected randomly. They became the subjects because they participated in field testing or elected to pursue OA from 1992 through 1994. The pre-1992 schools are those receiving accreditation in 1993 and 1994. They began their improvement processes between 1987 and 1991. The 1992 and 1993 schools are the entire cohort seeking approval for initiating OA candidacy those years. The 1994 sets of improvement goals are a sample of 84 out of the 291

improvement plans submitted that year. They come from one state in the NCA region.

Table 1 Yearly Breakdown Of OA Participation

	Number of Schools	Different States	Average Goals/Plan	Goals Submitted
Pre-1992	68	14	4.3	293
1992	165	17	4.7	772
1993	320	17	4.7	149
1994	84	1	4.3	360
TOTAL	637	20*	4.5	2921

* *In all twenty states had schools seeking candidacy during this study, though several states participated some years and not others.*

The schools varied by size, location, staffing ratios, education levels, and governance. Included in the sample are 19 different states and one overseas Department of Defense region, with levels of schooling spanning from pre-school through 12th grade, differing school sizes and staffing ratios, and public and private governance. Forty five percent were high schools, 23% middle or junior high schools, and 32% elementary schools. Enrollment averaged 710 students, and staffing 49 full time professional employees (FTE). Compared to figures for U.S. schools in 1993 (Department of Educational Statistics 1993), the average enrollment was 40% higher and average FTE was 47% greater. However, the student-to-staff ratio of 14.6 is close to the national average of 14.7. The ratio of private-to-public schools (8 to 92) also matched that of an average of the states from which these schools came.

Peer Reviewers

The peer reviewers include 245 people who took part in the first three years of OA candidacy peer review: 45 in 1992, 92 in 1993, and 103 in 1994. Random sampling was not used for peer reviewer selection. Almost three-quarters of the reviewers (73%) were chosen by NCA state directors. The remainder (27%) were walk-in volunteers who attended the annual conference in Chicago. Eighty-four percent of all reviewers currently were involved in OA processes in their schools or preparing to seek OA within the following year. Seventy-nine percent of all reviewers reported that they had no experience with the OA peer review process before coming to Chicago. In 1994, 33% of reviewers said they had used the holistic decision for peer review in their home states.

Reviewers most often were employed in schools. The majority worked as school principals (63%), and the remainder came from district offices (24%), institutions of higher learning (8%), or school boards and state departments of education (2%). Forty-eight percent of all reviewers worked in high schools or high school district offices. The rest were employed by K-12 (kindergarten through high school) districts (33%), elementary schools (13%), or junior high or middle schools (6%). Sixty percent of all reviewers in 1992, and 50% in 1993 and 1994 were male.

Data Collection

The materials used in this study for candidacy application, peer review, and reviewer surveys were linked closely in format and content. The review document, "Review Process for Schools Seeking OA Candidacy" was developed from criteria for compliance with OA found on the inside cover of the "Review Process for Schools Seeking OA Candidacy". To maintain consistency, the reviewer surveys

were constructed to reflect the holistic decision and diagnostics in the "Review Process for Schools Seeking OA Candidacy. In all, four documents were used as sources for data including:

- "The Report for Outcomes Accreditation Schools" submitted by all OA candidacy applicants
- Reviewer registration sheets for the OA candidacy review session in Chicago
- The "Review Process for Schools Seeking OA Candidacy" sheet which peer reviewers filled out for every school they evaluated.
- The "NCA Reviewer Surveys"

School Data

The main source of information about schools was supplied by the schools themselves in the" Report for Outcomes Accreditation Schools" document. On this form schools reported their city, state, school enrollment, the number of professionals employed (FTE), school level (grades included), and school governance (public or non-public) on this form. Also included are the school's improvement plan and a description of the steps the school took leading up to the development of the improvement plan (commitment; resources assistance, school profile study, selection of target areas based on data about student performance), and the school's set of improvement goals.

Peer Reviewer Information

Background information on the peer reviewers was obtained from the registration records for the annual regional OA candidacy review and from surveys returned by reviewers. Reviewers were assigned numbers at the time of registration that they recorded on each review form. These numbers were used to link individual reviews to a specific reviewer. The school OA candidacy reviews

and reviewer information for 1992, 1993, and 1994 also were collected at the annual regional review in Chicago and are complete.

The peer reviewers were surveyed for general perceptions of OA processes and OA precandidacy school efforts. All reviewers received the "NCA Reviewer Survey" from 1992 through 1994. In 1992 these surveys were mailed to each participating reviewer; 61% were returned. In 1993 and 1994 reviewers were given surveys which were collected from them before they left the candidacy review. The 94% of the surveys were returned in 1993 and 86% in 1994.

Measures

Peer Review Based On A Four-Point Forced Choice Scale

Admission to candidacy from 1992 through 1994 was granted when two peer reviews of the school's set of improvement goals had average ratings of 2.5 or more. Ratings are assigned by individual reviewers in response to the holistic decision question, "If a school improvement plan having this set of target area goals were faithfully implemented and learner outcomes were noticeably enhanced in these areas, to what extent would overall learner outcomes levels in the school have improved" (NCA peer review form). In addition, reviewers rated the school's plan on the seven diagnostics that are criteria for OA candidacy using the four-point scale. Some also added written comments to the school about their plan.

During a review of school materials the reviewer had to choose a rating level that implied the set of goals were either clearly acceptable (a rating of three or four) or unacceptable (a rating of one or two). In addition, because four rating levels were used, the reviewer indicated the degree of acceptability. In the peer review a rating of 4 signaled the goals were "Exemplary," 3 "Acceptable," 2 "Not Quite Acceptable," and 1 "Unacceptable." A rating of 4 indicates the reviewer

believes the set of goals will have a significant impact in improving learner outcomes. On the other hand, a rating of 1 means the reviewer thinks the school's goals will have no significant impact on improving learning in the school. A 3 denotes the reviewer's feeling that improvement will be fully acceptable, and a 2 that the school's plan will not make enough impact on student learning.

In order to compare the schools' sets of goals to the holistic decisions they received from peer reviewers, it was necessary to build a means of measuring the degree that the set of goals met criteria for OA candidacy. To do this, contents of OA candidate goal sets were analyzed using a coding system developed from the diagnostic feedback found in the peer review form. The coding categories employed for content analysis were established through a survey of school improvement plans submitted in 1992 and 1993 (Flanders 1993). The coding process was developed, tested, and refined over a two-year period in a variety of settings with educators from the NCA region (Wick and Gose, 1994).

Since NCA used measurability, higher level skills, level of learning, focus on student learning, and equity as indicators of improved quality, they also were used as quality indicators in our content analysis. The coding for each goal included the type of expected outcome accomplishment(complex behavior, indicator, implementation of a process, or organizational), the curricular or extra-curricular target selected, notations for evidence of decision making based on student data and equitable expectations for improvement, and the level of learning being addressed (skill and knowledge acquisition, or the integration and use of knowledge and skills in complex activities. The specific curricular or extra-curricular area targeted for improvement in each goal also was identified. Exemplars of the coding can be found in Appendix I.

The match to template (MTT) was constructed to provide an overall rating, roughly approximating a standardized holistic decision for the purpose of assessing the degree that contents of improvement plans matched OA criteria. The MTT was based on several assumptions. First, the presence or absence of MTT can be identified through constructs for each of the OA criteria. Second, the percentage of goals that had indicators for constructs of each diagnostic category could be averaged together to give a rough approximation of the extent that the school's goals meet OA criteria. These constructs, or characteristics, included the goals stated as student learning outcomes, equity, selection based on student data, challenge to the school's students, focus on higher level skills, involvement of the school's staff, and coordination and integration as a set. Finally, the MTT of a school's plan converted to a four-point scale can be compared to reviewers' holistic decisions because they presumably measure the same things.

An example of the construction of a MTT for a fictional school improvement plan can be found in Appendix II. Four steps were required to determine the extent to which a set of school goals met OA criteria based on the contents of the school's goals. These were a content analysis of the goals in a school's OA plan to identify OA criteria in each goal statement, a determination of the percentage of goals in the plan that had each criteria, an overall average for compliance from the percentage of each criteria for the school's set of goals, and conversion of the overall average from a 100-point scale to a four-point scale as used in the rating for the holistic decision and diagnostic criteria.

Procedures

"The Report for Outcomes Accreditation Schools" files were collected on site during the 1992 and 1993 reviews for OA candidacy in Chicago. The data

obtained from them is fairly complete. The same is true of the information taken from the final reports submitted by the 1993 and 1994 OA schools receiving full accreditation. In 1994, a copy of all of the candidate goals for one state were obtained from that state's NCA office.

Peer review sessions in Chicago began with a short lecture on OA criteria, models of goals and ratings, and a practice review. This was done to calibrate reviewers' holistic decisions before they began the actual reviews of school materials. Each improvement goal set received two independent reviews. Ratings were assigned by individual reviewers on the review form in response to the "Holistic Decision" question. The average of both reviews on the holistic decision determined admission to OA candidacy for the school.

Two separate databases were developed to answer the research questions that are the focus of this study. One linked the contents of school improvement plans and school characteristics to the OA template using the MTT. Descriptive and comparative analyses were made on the basis of several factors including year (pre-1992, 1992, 1993 and 1994), state, governance, school level, school size and staffing ratios. Changes in the difference between MTT and holistic decisions were investigated through year-to-year comparisons.

The other database linked the MTT to school characteristics, peer review ratings, and reviewer characteristics. The accuracy of ratings and peer reviewers were then compared on the basis of reviewer characteristics. Reviewer accuracy was identified using an average of the absolute difference between the reviewer's ratings and the average MTT of the school plans he or she reviewed. Correlations also were calculated on the peer reviewers' personal and averaged holistic decisions and the MTT of the 1992, 1993, and 1994 candidates that they reviewed.

Finally, regression equations were employed in an attempt to describe influences behind variations in the rate that schools from different states integrated OA criteria. Similar steps were taken to describe the development of accuracy in peer evaluation and its relationship to on environmental or individual factors.

Findings and Discussion

There are several findings relating to OA change that are measured by the assimilation of characteristics in candidacy preparation. First, school improvement plans used during the term of this study demonstrated an improved match to the criteria. Second, the use of separate OA criteria developed at different times, and rates in different states. Third, there was little difference between the average of reviewer ratings for admission to candidacy from year-to-year. However, the gap between the reviewer's holistic decisions and the MTT of the candidates he or she reviewed grew smaller each year. Environmental pressure and support to change, school autonomy to deal with changing, the degree criteria required a departure from conventional practice, and individual and organizational activism were significant in affecting the rate with which schools and reviewers used OA guidelines for school improvement.

Did The Schools And Peer Reviewers Learn To Use The Criteria?

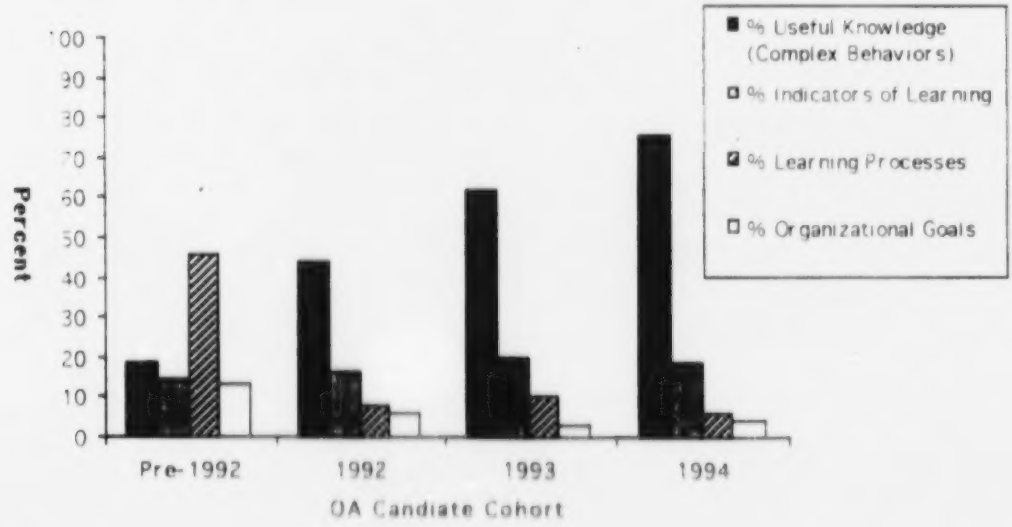
Schools

The targets schools selected for improvement varied little from year-to-year. The majority of goals examined targeted Language Arts, values related to learning, or tools and processes that support learning. From the pre-1992 group through the 1993 cohort, 25% of all goals targeted language arts, another 28% targeted student values related to learning, and 25% targeted processes and tools that support

made by the 1994 sample. The percentage targeting language arts, values that support learning, or tools and processes that support learning in 1994 varied no more than 3% from those for pre-1992 through 1993 schools.

However, compliance with OA guidelines is not reflected in target selection. It can be seen in the type of outcome expected and from evidence that individual criteria was being addressed. Compliance was low for the pre-1992 and 1992 cohorts, but increased in the following two years (see Figure 1). Prior to 1992, only 11% of all goals stated what students would do (complex behaviors) as a result of improvement efforts. In contrast, 75% of the 1994 sample's goals had expectations for improved complex behaviors. The previous year, only 54% of all goals from this state were complex behaviors. Overall, indicators remained the only persistent feature of expectations other than complex behaviors throughout this investigation.

Figure 1 Percent Of All Goals By Type Of Outcome, Pre-1992 - 1994



Change was observed from year to year in the integration of OA criteria. Prior to 1992, goal statements in school improvement plans exhibited all of the criteria about 12% of the time. This improved to 76% in 1994. The percentage of goals that addressed each criteria varied from year to year as overall use of the criteria grew (see Table 2). On this table it can be seen that focus on student learning was high from the start. By 1994 it was the most standardized feature of goals. The use of curricular integration and higher level skills improved most for the 1992 cohort and measurability for the 1993 cohort. On the other hand, equity was the last criteria to be assimilated. By 1994, equity was evident in 84% of the goal statements from the sample state. The previous year only 54% of all goal statements from this state had equitable expectations.

Table 2 Goals Addressing Major Aspects Of The OA Template

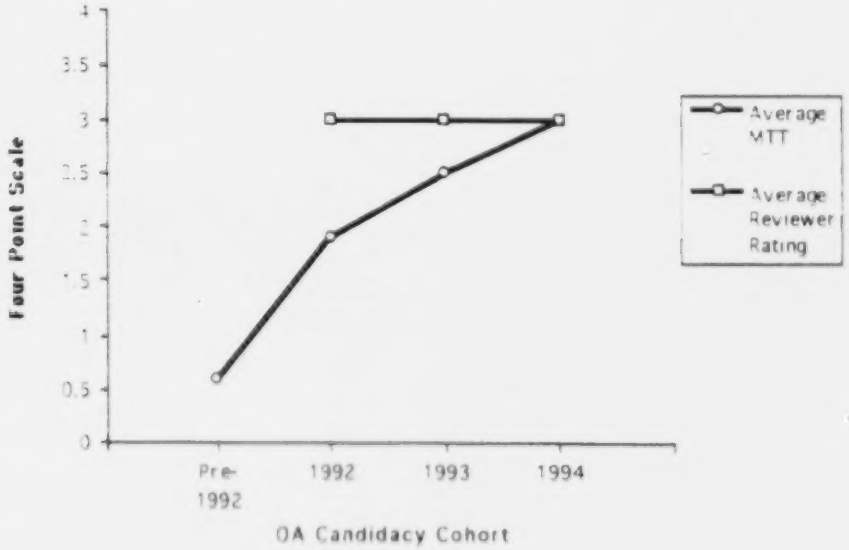
	Pre-1992	1992	1993	1994
Focused on student learning	67%	66%	91%	97%
Demonstrates measurability	52%	57%	77%	78%
Integrates Curricular Areas	49%	75%	76%	71%
Targets improvement equitably	36%	39%	48%	86%
Targets higher level skills	38%	57%	71%	76%

Most growth in addressing the OA template in goals statements occurred in those areas that were the least (equity and higher level skills) or most (student learning outcomes) developed prior to the peer review.

Peer Reviewers' Holistic Decisions

During the first two years of the peer review, false positive decisions were predominant because peer ratings consistently exceeded the MTT. Averaged peer review ratings on the holistic decision varied insignificantly from 1992 through 1994 [F (2, 565) = .454, p > .05]. However stable the ratings, the gap between holistic decisions and the degree to which sets of school goals addressed OA criteria closed (see Figure 2). The average 1992 reviewers were more than one point apart from the degree of OA criteria evident in the candidate's set of goals. By 1993 the gap between the two closed to less than one-half point. By 1994, almost 75% of all reviewers differed less than one half a point, on average, in their holistic decision ratings from the MTT of the schools they reviewed.

Figure 2 Averaged Holistic Decision Ratings And The Degree Of Match To Template (MTT) Of Sets Of School Improvement Goals



Reviewers' holistic decisions were stable. Still, the difference between holistic decisions and the MTTs decreased between 1992 and 1994.

The increasing accuracy of individual reviewer's decisions provided further evidence of the growing agreement between decisions and the content of school improvement plans. Accuracy improved each year (Table 3). By 1994 75% of all reviewers, on average, differed less than one-half point from the MTT of the schools they reviewed. In comparison, during the review in 1992 at the NCA annual meeting less than one quarter of all reviewers were as accurate, and in 1993, only half met the same level of agreement between decisions and contents of goal sets.

Table 3 Yearly Holistic Decision Rating Accuracy* On A Four Point Scale

	1992	1993	1994
Fourth Quartile	0.0 - 0.8	0.0 - 0.2	0.0 - 0.05
Third Quartile	0.9 - 1.4	0.3 - 0.5	0.25 - 0.3
Second Quartile	1.5 - 1.6	0.6 - 0.75	0.35 - 0.4
First Quartile	1.7 - 2.5	0.8 - 2.12	0.5 - 1.4

* Accuracy is measured by finding the difference between the MTT for the school's set of goals and the reviewer's holistic decision rating. The most accurate HDs are found in the fourth Quartile; the least in the first.
On the average the first quartile in 1994 is at least as accurate in making holistic decisions that agree with the OA template as the third quartile was in 1992.

How Did Change In Use Of The Criteria Develop?

Schools

Characteristics or environmental factors were compared to the school's selection of targets or identification of expectations to determine their influence on the later. These comparisons were made on the basis of the year of attempted candidacy, schooling level, enrollment, staffing ratios, governance, and state. Correlations between percentages of target areas selected and outcomes indicated for improvement were strong (see Table 4) regardless of year, level, enrollment

staffing ratios, and governance. However, correlations between the percentage of each expectation and states were weak and sometimes negative.

Table 4 Correlations for target area selections and expectations for improvement

	Targets selected ($df=91$)	Expectations identified ($df=17$)
LEVELS		
Elementary / Jr. High	0.90**	0.91**
Elementary/High School	0.93**	0.93**
Jr. High/Middle School	0.97**	0.98**
ENROLLMENT QUARTILES [1]		
first and second	0.94**	0.95**
first and third	0.94**	0.95**
first and fourth	0.95**	0.93**
second and third	0.96**	0.95**
second and fourth	0.95**	0.95**
third and fourth	0.95**	0.95**
STUDENTS/FTE QUARTILES [2]		
first and second	0.94**	0.97**
first and third	0.87**	0.96**
first and fourth	0.80**	0.91**
second and third	0.95**	0.97**
second and fourth	0.94**	0.94**
third and fourth	0.96**	0.96**
GOVERNANCE		
Public and nonpublic	0.75*	0.97**

* significant at the .01 level for a nondirectional test.
** significant at the .001 level for a nondirectional test.
[1] Mean enrollments for each quartile are: first - 413 students; second - 583 students; third - 813 students; and fourth - 873 students
[2] Mean students per professional staff member (FTE) for each quartile are: first 10 students, second 12 students, third 16 students, and fourth 19 students

Further comparisons were made to determine if these differences were significant. No significant difference was found between states in percentage of target areas selected [F (19, 380) = 0.067, $p > .05$)]. However, when the percentage of outcomes indicated for each complex behavior, indicator of learning, implementation of learning processes, and organizational outcomes were compared by state they were significant [F (19, 200) = - 5.23 , $p < .001$]. Because some states submitted very few school plans (less than 10), they were eliminated and a

second analysis of variance was conducted on the remaining 14 states with larger groups of candidates. These findings were also significant for differences in expected learning outcomes [F (13, 140) = 7.49, $p < .001$], though, no significant difference was found for target selections [F (13, 247) = 0.039, $p > .05$].

Several things contributed to differences between states in their expectations of student learning improvement. The rate at which states integrated all the criteria (MTT) from year to year varied from a loss of 20% to a gain of 33%. Most losses were observed the first year criteria was set. This is probably due to two factors. First, the large increase in schools seeking candidacy in 1992 and the fact that sets of goals for pre-1992 schools were collected in 1993 and 1994. It is likely that the goals of pre-1992 candidates reflected some of the evolution of criteria in use since formalization in 1992. The average gain in integration of criteria was from pre-1992 to 1992 was 6.4%. The 1993 average gain improved to 14%, and the 1994 sample gained 15% over the previous year.

Between the pre-1992 schools and the 1992 schools, state-level improvement was observed most often in integrating higher level skills, followed by student learning outcomes and data-based decision making. However, evidence of progress in integrating equity into expectations was lowest in seven out of 12 states in 1992. Overall improvement in using student data for decision making was strong when the 1992 and 1993 state cohorts were compared (See Table 5). However, higher level skills showed the least improvement in six states and student learning outcomes in five.

States also had preferences for certain complex behaviors, indicators, processes, or goals without outcomes (see Table 6). From 1992 through 1993, some states always had outcomes for caring for self and others, though the same

outcomes occurred infrequently in the plans from other states. On the other hand, making, fixing and growing things, sciences, social sciences, and mathematics were a rarely expected outcomes in the plans from most states. Furthermore, indicators always were included in the plans of five states as were processes though never in others.

Table 5 Gains In Major OA Criteria By Number Of States

	Student learning outcomes	Equity	Data-based decisions	challenge	Higher level skills
Pre-1992 to 1992 (12 states)					
Most improved	3	1	3	0	5
Least improved	3	2	0	2	0
1992 to 1993 (16 states)					
Most improved	4	3	8	0	1
Least improved	5	3	1	1	6

Table 6 Number Of States By The Probability That A Specific Expectation Would Occur (In Order Of Frequency)

OUTCOMES	> 95%	75 - 95%	50 - 74%	25 - 49%	5 - 24%	< 5%
• Making, fixing, or growing things	-	-	-	-	4	15
• Using scientific or technical knowledge to explain and improve outcomes	-	-	-	2	6	11
• Using social sciences to explain interaction of humans, groups, and nations	-	1	-	2	7	8
• Using mathematical principles for decision making	-	-	-	5	7	7
• No expectations for learning outcome identified	5	-	2	5	2	5
• Applying problem solving and creative thinking skills	2	3	2	4	7	4
• Outcomes for schools, parents, or teachers (not directly related to student learning)	1	1	4	6	4	3
• Generating and receiving communication	2	4	7	4	-	1
• Responsibly caring for self and others	5	4	5	2	2	1
• Implementing a process that may improve learning	5	2	3	4	2	2
• Reaching indicators that student learning may have occurred	5	2	3	4	2	2

BEST COPY AVAILABLE

Peer Reviewers

It was discovered that the characteristics of school goals influenced holistic decisions differently from year to year. When regression was used to identify which criteria reviewers used in holistic decision making from 1992 through 1994 (Table 7), it was found that reviewers focused on some aspects of the OA template and shifted their interest away from or ignored others. In addition, despite the accrediting agency's recommendation that plans contain five goals, only the number of goals (four were the reviewer's preference) remained significant in all three years. However, the adjusted R^2 suggests that few of the characteristics had any significance for reviewers' decisions in 1992 and 1993. In 1994 it explained little more than 30% of the holistic ratings that year.

Table 7 Impact Of School's Sets Of Improvement Goals On Averaged Holistic Decisions

	REGRESSION COEFFICIENTS (*t* ratios in parenthesis)		
VARIABLES	1992	1993	1994
Total number of goals	-0.107	-0.079	-0.397
	(-2.227*)	(-1.940*)	(-3.961**)
Goals for different curricular areas	0.018	0.093	0.308
	(0.432)	(2.957**)	(2.825**)
Goals that indicate learning	0.096	0.023	0.036
	(2.250*)	(0.648)	(0.451)
Goals for learning processes	0.002	0.047	-0.069
	(0.041)	(1.127)	(-0.677)
Goals for complex behaviors	0.017	0.005	0.058
	(0.276)	(0.207)	(0.641)
Goals for school outcomes	-0.021	-0.263	-0.036
	(-0.027)	(-3.123**)	(-0.211)
Goals integrating learning	0.026	0.034	0.016
	(0.543)	(1.205)	(0.212)
Goals that can be measured	0.011	-0.026	0.081
	(0.246)	(-0.916)	(0.866)
Goals including all students	0.103	0.035	0.302
	(2.367*)	(1.647)	(3.404**)
Intercept	3.160	2.821	2.254
	(24.522)	(24.982)	(6.055)

BEST COPY AVAILABLE

Table 7, continued

Adjusted R^2	.007	.008	.032
Standard Error	.56	.52	.56
Sample size	164	310	84

* $p < .05$, for two-tailed test ** $p < .01$ for a two-tailed test

Since reviewer accuracy is the basis for reliability in peer review ratings, it was examined to see if accurate and inaccurate reviewers made decisions differently. Regression equations were employed to determine which OA criteria influenced inaccurate and accurate decisions in the first and fourth quartiles by averaged rating in the 1993 cohort (see Table 8). One hundred sixteen of the peer reviews were considered accurate because their averaged holistic decisions matched the averaged MTT of the candidates they reviewed. Fifty-one were identified inaccurate because the averaged MTT and holistic decision ratings were two to three points apart on a four-point scale.

The use of OA criteria for equity seems crucial. It was a significant difference in decision making leading to rating assignment between accurate and inaccurate reviewers. The accurate reviewers assigned significantly higher holistic decisions to goal sets with greater percentages of challenge and use of higher level skills. Inaccurate reviewers lowered ratings in response to higher percentages of equity in goal sets though, like accurate reviewers, they raised them for higher level skill use.

Use of OA criteria in sets of goals explained about 80% of accurate reviews in the top and bottom quartiles by averaged review in 1993. Only 42% of the decisions made by inaccurate reviewers could be accounted for by OA criteria with a high standard of error. It is evident that accurate reviewers in 1993 used most of

the criteria in rating assignments. The construct for challenge includes the average percentage of student learning focus, measurability, and curricular integration in goal sets. The construct for higher level skill use includes the average percentage of integration and equity. However, inaccurate reviewers sometimes rejected the OA concept of equitable improvement for all students when making their decisions.

Table 8 Influence Of School Improvement Goal Sets On Averaged Decisions From 1993 Fourth And First Quartiles

	Regression coefficients (*t* ratios in parenthesis)	
Variables	Accurate	Inaccurate
Percentage of goals that are student learning outcomes	0.074 (1.55)	0.511 (1.61)
Percentage of goals that are equitable	0.082 (0.805)	-0.594 (-1.828*)
Percentage of goals that are selected from student data	0.028 (0.313)	-0.275 (-0.957)
Percentage of goals that are challenging to the school's students	0.354 (1.704*)	-0.234 (-0.312)
Percentage of goals that are outcomes for higher level skill use	0.147 (2.15**)	1.126 (2.02**)
Percentage of goals that involve school staff	0.153 (0.906)	0.291 (0.476)
Percentage of goals that are coordinated and integrated	0.068 (0.331)	-1.379 (-1.526)
Adjusted R^2	0.80	0.42
Standard Error	0.33	0.89
Sample size	115	51

* $p < .10$, for two-tailed test ** $p < .05$, for two-tailed test *** $p < .01$ for a two-tailed test

Accurate reviewers assigned significantly higher holistic decisions to goal sets with greater percentages of challenge and use of higher level skills. Inaccurate reviewers lowered ratings in response to higher percentages of equity in goal sets though, like accurate reviewers, they raised them for higher level skill use.

DID SCHOOL CHARACTERISTICS OR FACTORS IN THE SCHOOL'S ENVIRONMENT FACILITATE OR CONSTRAIN THE RATE OF CHANGE?

Improving Or Slowing The Rate Of Schools' Progress Toward Integrating Criteria

Several factors affected the rate that schools integrated OA criteria (see Table 9). Most significant for increasing the rate of criteria use was a state testing program balanced with a healthy degree of school autonomy. The years since formalization of OA criteria in the peer review and improvement in equity were also significant factors for the rate of change. The adjusted R^2 suggests that the combined factors account for 46% of the degree of accuracy between reviewer's average HDs and the MTT of the candidates they reviewed.

The sixty-six percent of improvement (adjusted R^2 in Table 9) by state cohorts in use OA criteria can be explained by a healthy degree of school freedom from regulatory stress tempered by some pressure from state testing programs. Interestingly, the degree of school autonomy (the percentage of freedom from four constraints - state testing, performance reporting, mandated school reform, and goal setting by the school's superordinate ordinate agencies) was most significant for facilitating the rate of state's criteria integration. However, superordinate agency pressure also served to facilitated the rate as seen in significant improvements associated with state testing programs. These tests motivate and focus schools on improving student learning, therefore, they align closely with the intent behind OA criteria.

Time and the difficulty of using criteria also were factors for learning or "getting it." The passage of time is evident in the predicted values of years the peer review had been used. An increase in the percentage of equity indicated less improvement in MTT over the previous year. Since equity was usually last of the criteria used in goal sets most of the work on using the other six criteria was

BEST COPY AVAILABLE

already accomplished. So, state cohorts "got it" in increments usually leaving the most difficult criteria - equity - until the end.

Peer Reviewer Accuracy

Reviewer accuracy was examined to see if accurate and inaccurate reviewers made decisions differently. None of the variables relating to the reviewer's state, previous experience in making the holistic decision, or employment background made a difference in accuracy. Multiple regression analysis was conducted on the holistic decisions of 200 reviewers to see what contributed to improved accuracy (see Table 9). Certain factors improved were significant for improved accuracy. These were: (1) state engagement in OA, (2) individual engagement by the reviewer in OA learning, and (3) the extent to which candidates' sets of goals addressed the template.

Finally, reviewer accuracy was investigated through regression to see if it could be identified by characteristic of reviewers or ecological influences. The decisions of 200 reviewers from 1992 through 1994 were examined (see Table 10). Forty-five 1994 reviewers were omitted because OA improvement plans provided only a sample of the entire cohort. Therefore, no information was available on the accuracy of these reviewer's decisions.

Regression explained 46% of accuracy (see the R^2 in Table 10). An increase in accuracy is indicated in Table 10 by a negative ratio because the difference between the average MTT of school plans and reviewer's reviews of those plans decreases as accuracy increases. Several significant factors in the development of reviewer's accuracy are indicated. Reviewers learned most rapidly when they exercised initiative by taking advantage of a variety of the opportunities -- formal and informal contacts with others involved in OA, on-site school observations, seminars and workshops, consultation for state offices, and the use

of OA print and video. Improvement in candidates' use of criteria in their plans was most significant, and was far more likely in 1994 than in 1992.

Table 9 Regression For Improvement In MIT As Defined By The Difference Between The Average MIT Of State Cohorts

Variables	Coefficient (*t* ratio in parenthesis)	Predicted value
Environmental Pressure On Schools To Improve Learning		
State testing program [a]	0.330 (2.533)*	0.02
School achievement performance reported by State [b]	0.054 (0.895)	0.26
School improvement mandated by state [c]	0.103 (1.600)	0.13
Superordinate agency identifies some of the school improvement goals [d]	0.148 (1.364)	0.20
Superordinate agency sets all of the school improvement goals [e]	-0.067 (-0.416)	0.69
Degree of school autonomy	0.437 (3.838)**	0.002 *
NCA Regional Pressure		
Years OA peer review has been used	0.252 (2.45)*	0.03
NCA State Organization Capacity To Assist Schools		
Years state has been involved in OA [f]	-0.13 (-1.84)	0.09
Percentage of the states' goals focused on student learning the previous year	-0.10 (-1.12)	0.29
Percentage of states' goals that were equitable the previous year	-0.246 (-2.52)*	0.03
Percentage increase in OA caseload from the previous year	-0.186 (-1.44)	0.18
Intercept	-0.239 (-1.89)	
Adjusted R^2	0.66	
Standard Error	0.09	
Sample size	24	

* $p < .05$, for two-tailed test ** $p < .01$ for a two-tailed test

a A dummy variable indicating the state has a testing program from US Department of Educational Statistics (DES) 1993 and 1996, and Cetron and Gayle 1992.

b A dummy variable indicating the state reports school performance based on student achievement found in DES 1993 and 1996, and Cetron and Gayle 1992.

c A dummy variable indicating the state has mandates for school improvement found in DES 1993 and 1996, and Cetron and Gayle 1992.

Table 9, continued

d A dummy variable indicating the school's goals are determined in part by an agency and in part by the school.

e A dummy variable indicating the school's goals entirely determined by a superordinate agency without the school's involvement.

f The percentage of freedom from four constraints - state testing, performance reporting, mandated school reform, and goal setting by the school's superordinate ordinate agencies.

g A count of the years since the beginning of peer review.

Table 10 Regression For Accuracy# In Decision Making, 1992 - 1994

Variables	Coefficient	*t* ratio
State OA Activity Level		
Previous reviewers[a]	0.002	1.686
Previous candidates[b]	0.0004	1.175
Reviewers present[c]	-0.004	-2.543**
Invited Reviewer[d]	-0.032	-1.973*
Average School MTT	-0.298	-4.865**
Reviewer Engagement		
Seeking OA Candidacy[e]	0.004	0.197
OA Workshop[f]	-0.423	-1.981*
NCA OA print and media use[g]	-0.071	-2.770**
OA Personal contacts[h]	-0.080	-2.686**
Years Holistic Decision In Use	-0.021	-1.201
Adjusted R^2	0.46	
Standard Error	0.11	
Sample size	200	

* $p \leq .05$, $-1.960 \leq t \geq 1.960$, two-tailed test ** $p \leq .01$, $-2.576 \leq t \geq 2.576$, two-tailed test

 # A negative result indicates a decrease in the difference between the MTT of the candidates reviewed and peer reviewers' holistic decisions.

a The percentage of peer reviewers coming from the reviewer's state prior to the current year, converted to a four-point scale.

b The percentage of OA candidates coming from the reviewer's state prior to the current year, converted to a four-point scale.

c The percentage of peer reviewers coming from the reviewer's state this year, converted to a four-point scale.

Table 10 continued

d A dummy variable signifying the reviewer was invited to participate by the state office.
e A dummy variable indicating the reviewer reported seeking OA candidacy.
f A dummy variable indicating the reviewer reported OA workshop attendance.
g A count of OA print or video sources reviewer reported using.
h The number of the different types of personal contacts (state office, school visits, informal peer discussions, and conversations with peers seeking specific information) the reviewer reported.
i State averages of the reviewer's cohort were assigned to 35 reviewers who did not return surveys. In addition, the reviewers who did not conduct reviews on plans from the sample state in 1994 were excluded.

Conclusions

Using behaviors linked to the use of the criteria for admission to OA candidacy required learning. We conclude that schools and peer reviewers learned because they demonstrated behaviors that were not developed at the beginning of this study. In 1992, schools could not write goal statements reflecting the improvement of learning with equity, nor could individuals accurately judge if goals addressed these criteria. In 1994 goal sets and peer evaluations reflected application of the criteria.

The Rate Learning Progressed

The most important ingredient for improved use of the criteria by schools and by reviewers was time. Time was required to accommodate personal and organizational learning. Considering this is just one step in the accreditation agency's process for improving learner outcomes in schools it seems that the time required for successfully achieving improved learner outcomes extends beyond the scope of this study.

Based on our observations, it is our opinion that the time required for learning will have the greatest impact on participants involved in the first years of implementation of OA in schools and classrooms. The challenges they will face

include learning how to develop strategies to achieve their goals, establishing baselines for improvement, implementing improvement plans in classroom teaching and learning, evaluating their effectiveness, adjusting efforts to stay focused on their goals, and documenting authentic improvement. Therefore, we also believe that the time their successive cohorts will take to these things proficiently will shorten because of the evidence that the peer organization and NCA state organizations benefited from learning. Fortunately, we observed that knowledge about the use of OA criteria was conserved by the accrediting agency and its members to be passed on to the next cohort resulting in the reduction of replicated efforts and stagnation.

In addition, the rate of change was sensitive to several ecological factors that, on first glance, are not directly linked to OA processes. Beliefs that aligned with OA criteria at initiation made a difference in the rate schools from different states used it. When states had expectations for testing outcomes, indicators commonly were used by the schools for goals and reviewers accepted them in improvement plans despite issues with equity. If the state set other priorities for improvement these were also reflected in school expectations and reviewer accuracy.

Some of the factors that influenced that rate of change in states seemed contradictory, for instance the significance of both state testing programs and school autonomy. This may implies that some external pressure to improve student learning at the school is important ecological direction reinforcing OA. However, overwhelming pressure on the school - external mandates for reform, school performance reporting, and identification of all the goals for improvement in addition to testing - slowed the schools' ability to focus on improvement of student learning defined in OA criteria.

All this suggests that change can be accelerated or decelerated depending on what organizations and persons face beyond the scope of the change. Change makers do need to be sensitive to this or they risk finding themselves defeated by unexpected delays or resistance. In the future, it will be important to determine the potential of ecological factors and personal or organizational characteristics for sustaining OA change. Their influence could be investigated through growth, stagnation, or extinction in different school settings and states.

In addition, we believe that the theories about school improvement that formed the base for OA will shift and be reshaped by external influences. Agencies with systemic connections to the schools participating in OA will likely have a part in influencing the direction of this shift. One example of redefinition was seen during this study in the accrediting agency's response to criticism about the inclusion of values. In even though most schools selected affective goals for values that supported students' learning, in 1994 they dropped the recommendation for two affective targets. Another change that may influence the substance of targets or expectations is the activity begun in most states under federal inducement to formalize common standards for learning. So, observing the interaction of common state standards with OA processes will be of particular interest.

Missing from this investigation is important information the on roles of school and school district culture in the use of OA. Further research in this area would complete a picture only has begun to be sketched by this study. We believe this information could help predict the rate of school adaptation to OA and the development of reviewer accuracy with greater accuracy.

However, there were some clues to the importance of school communities and school districts in OA development, though no way was available to us for collecting this data. Most notable were reports in OA documents of school

professionals' skepticism and resistance to this process and evidence of tight control by school districts over improvement plans. In some cases school professionals were kept at arms length from their own school's improvement efforts. School districts would, in different situations, identify their schools' goals, use central office personnel to interpret the data, and produce school studies or student profiles in an assembly-line manner.

How Learning Grew

Difficulty with specific criteria influenced school capacity to change. The five major OA criteria can be listed in order of difficulty based on year to year comparisons of gains. For these schools and NCA state organizations they were: (1) focus on student learning, (2) tie between the use of data relating to student learning to make decisions based on student data and the use of higher level skills, (3) integration of learning through the use of expectations for complex behaviors, and (4) expectations that learning would improve for all of the school's students or equity.

Expected outcomes for learning improvement varied because schools in some states integrated criteria at different rates and integrated different criteria in different years. Essentially, states already had or did not have a degree of "native" knowledge that aligned with the criteria for OA. Most states' schools were good at focusing on student learning from the very start, if not then that was the criteria they improved in using at first. However, some were also good at using measurement, higher level skills, or integrating knowledge from the onset. We also noted that about half of the states improved most in the criteria that they had developed the least the previous year; and least in the one that had been most

developed. In addition, half of the states also improved the most in using criteria that received the lowest average on diagnostic feedback rating the previous year.

Equity was a difficult concept for states, schools and reviewers. OA criteria specifically requires that improvement include equitable expectations for student outcomes. We believe that field educators had difficulty with this concept because they had to shift a customary definition of equity used in practice to fit the one found in OA criteria. Essentially the concept of equity had to evolve from equal access to schools (even though different treatment and expectations were acceptable) to include common expectations for improvement in addition access. As a result of this shift in thinking, we feel that equity was being improved in these schools. In addition, between-school equity also improved because the school participants standardized their expectations for learning improvement to external criteria.

We also suggest that the participants in this shift were not sensitive to it. In addition, we think people may be unaware of this change. Primarily because the label for "equity" remained the same, despite a shift in beliefs that contributed to changing the accepted meaning of what was equitable. However, the underlying construct for beliefs about equity did change between 1992 and 1994 for reviewers and schools. In all probability the schools and educators participating in this study would be as adamant today as they were five or six years ago that they are firm believers and practitioners of equity.

Accuracy In Peer Review Followed the Use Of Criteria In Goal Statements

Accuracy in holistic decision making appears to follow the use of criteria in goal statements, not precede it. So, as OA criteria were used, expertise in evaluation developed. Assimilating OA criteria was crucial for reviewer accuracy

BEST COPY AVAILABLE

and took about two years to develop. Compliance with OA guidelines in the set of school improvement goals had the most effect on reviewers' accuracy supporting this hypothesis. The stability of the averaged peer review ratings also indicates that the accuracy of the reviewers may have kept pace with the development of knowledge about the use of criteria in goal statements.

Therefore, these judgments are accurate reflections of the understanding of the use of template criteria as far as it had developed at that point. The collective judgments of peer reviewers reflected the criteria that was usable at the time, so, schools weren't punished for not using something still under development. This leads us to believe that peer review based on holistic decisions is well adapted to the initial implementation of school improvement guided by loose criteria for school improvement.

In the first two years, false positive holistic decisions far outweighed false negative ones. Fortunately, false positive peer reviews did not deter development and use of the OA criteria. We tend to believe that the use of peer reviews may have been fairer than measurement through indicators. The MTT is an example of a rigid evaluation system based on indicators that could have been used to determine if schools were ready for improvement. If the MTT had been used to qualify candidates for school improvement, more than 50% of the schools would have been barred from candidacy in 1992.

However, irregularities between reviewer ratings, even if only by one point were the source of dissatisfaction among NCA state directors and school districts with the peer review. They appeared most often when schools in the same district or from the same state submitted identical goals and received different ratings. Submission of identical plans does not communicate the spirit of attempting school-

based improvement, but reviewers were not finely tuned, either. This resulted in complaints and a loss of confidence in the review process.

The Rate Of Criteria Use And Accuracy Of Peer Reviews Had Systemic Influences

In this investigation the rate of organizational use of OA criteria is related to the state NCA organizations' ability to support development of the use of criteria, and superordinate agencies pressing schools for improvement. Schools applied the criteria at accelerated rates when they came from state environments where opportunities to learn about OA were provided, and where there was some formal pressure to improve student learning in addition to organizational autonomy to attend to improvement. From the study of peer reviewers it could also be seen that persons learned most rapidly when certain conditions existed that provided them with the opportunity to learn and when they took the initiative.

Candidate's compliance with the template in the set of school improvement goals had the most impact on reviewers' holistic decisions. Since it took time to evolve application of the criteria in improvement plans this makes sense. Thus, the accuracy of peer reviews (the difference between reviewer holistic decisions and candidate match to template) increased as OA candidate goals integrated and adhered more closely to the template. Reviewers' engagement in acquiring knowledge about the process and NCA state office activity centered on OA also contribute to accuracy in holistic decision making.

Summary

School organizations and individual members of the accrediting organization exhibited learning behaviors. They progressed forward through incremental use of the criteria for OA. As use of criteria fully developed it was

integrated into goal statements and then into peer evaluations. Furthermore, the rate progress was influenced by several factors. These were: (1) time, (2) the capacity schools had already built up to deal with criteria, (3) reviewers' personal initiative for acquiring knowledge about OA, (4) the opportunities states could offer to individuals and organizations for learning and exchange, (5) a degree of organizational autonomy and directed by some external pressure on schools to specifically improve student learning.

REFERENCES

Bruer, John. *Schools for Thought: A Science of Learning in the Classroom.* Cambridge: MIT press, 1993.

Darling- Hammond, Linda, and Milbrey McLaughlin. Policies That Support Professional Development in an Era Of Reform. *Phi Delta Kappan*., April 1995, pp. 597-604.

DeBoard, Robert. *The Psychoanalysis of Organizations: A Psychoanalytic Approach to Behavior in Groups and Organizations*. London: Routledge 1978.

Department of Educational Statistics (DES) *The Digest of Education Statistics*. Washington, D.C.: US Department of Education: Office of Educational Research and Improvement, 1993.

Flanders, Anne. *When Peers Rate School Improvement Plans*. Tempe: North Central Association Commission on Schools, 1993.

Kanter, Rosabeth, Barry Stein and Todd Jick. *The Challenge of Organizational Change: How Companies Experience It and Leaders Guide IT*. New York: Free Press 1992.

Knapp, Michael, Patrick Shields, and Brenda Turnbull. Academic Challenge in High Poverty Classrooms. *Phi Delta Kappan*, June 1995, pp. 770-777.

Lewin, Kurt. *Field Theory in Social Science*. New York: Harper, 1951.

Louis Karen and Miles. *Improving the Urban High School. What Works and Why*. New York: Columbia Teacher's College, 1990.

McLaughlin, Milbrey. Learning From Experience: Lessons From Policy Implementation. *Evaluation and Policy Analysis*, #9 1987, pp. 171-178.

_____. "Discussant Remarks: Understanding State Efforts to Improve Teaching and Learning" a session held at the annual meeting for the American Educational Research Association, New Orleans, April 1994.

Meier, Deborah. Can the Odds Be Changed? *Phi Delta Kappan*., January 1998, pp. 358-360.

Moffett, James *The Universal School House: Spiritual Awakening Through Education*. San Francisco, CA: Jossey-Bass, 1994.

North Central Association, *A Focus on Student Success: A Handbook for Schools Seeking Outcomes Accreditation* . Tempe: NCA Commission on Schools 1994.

Ross, Lee and Richard Nisbett, *The Person and the Situation: Perspectives of Social Psychology*. New York, McGraw-Hill, 1991.

Rothman, Robert, *Measuring Up: Standards, Assessment, and School Reform*. (San Francisco: Jossey-Bass 1995).

Schlechty, Phillip. *Inventing Better Schools: An Action Plan for Educational Reform*. San Francisco: Jossey-Bass, 1997

Senge, Peter. *The Fifth Discipline : The Art and Practice of the Learning Organization*. New York: Doubleday Currency, 1990.

Tyack, David and Larry Cuban. *Tinkering Toward Utopia: A Century of Public School Reform*. Cambridge: Harvard University Press, 1995.

Wagner, Tony. *How Schools Change: Lessons from Three Communities*. Boston: Beacon Press, 1993.

Wick, John and Thomas Sarterfiel. Indicators, Criteria, And Assessment: The Great American School. A paper presented to the Great American School Conference:Breckenridge: June 1992

Wick, John and Ken Gose. *Improving Student Performance in Your School*. (Dubuque, IA: Kendall/Hunt 1994).

Appendix I

Improvement Target Areas And Exemplars Taken From OA School Improvement Plans

GOAL CATEGORIZATION BY LEARNING TARGETS

Following are examples of several goals classified by their primary academic target.

- Language arts:
  "Students will demonstrate ability to comprehend the main and subordinate ideas in written work and select appropriate ways to communicate these ideas."
- Math:
  "Students will solve complex math problems involving several steps and operations."
- Problem-solving and critical thinking skills:
  "Students will improve their ability to solve, analyze, and evaluate problems using logical steps and appropriate resources."
- Study skills:
  "Students will assume responsibility for their learning through preparation, use of timelines, application of inquiry skills, cooperation with others in group tasks, and the development of personal standards for the quality of their work."
- Social sciences:
  "Children will share knowledge of their heritage, language, culture, and life experiences."

CODING FOR QUALITY AND EQUITY

Expectations for Outcomes

For the purpose of this study, a target-area goal was considered a student learning outcome (SLO) if it focused on skills or knowledge used by the students. Therefore, complex behaviors, indicators, and processes were all considered SLOs. Goals that were either unclear or focused on targets unrelated to student learning were not SLOs. Following are three examples of goals from school improvement plans that can be classified as student learning outcomes:

- Complex behavior:

"Students will increase their knowledge and ability to use various technology as tools for learning, information management, and communication."

- Indicator of learning:
  "Reading comprehension scores will improve."

- Implementation of a process supporting learning
  "Students will make portfolios in preparation for application to higher education or future employment."

Goals that do not address the use of knowledge and skills include indicators, processes, outcomes for parents, teachers, and organizations. Student learning goals for two different complex behavior strands follow.

- Communication:
  "In all subjects students will demonstrate the ability to produce written documents in standard English (mechanics, usage, spelling)"

- Caring for self and others:
  "Students will make responsible choices for their social and emotional health".

An indicator targets a narrow outcome for improvement through test scores, grades, or change in a count of student behaviors (absenteeism, library material checkout, discipline notices, counselor contacts, survey responses, teacher observations). Indicators do not provide in-depth information about how useful this is for students, but they usually imply which complex behavior the school wishes to improve. Two examples of indicators are: "Scores on weekly spelling tests will improve." and "Students will be able to restate feelings and responses in non-judgmental terms."

Process goals are related to student learning, but they fall short of describing students' use of complex behaviors. Like indicators, they are an important part of school improvement processes but are not precise targets for students' use of knowledge and skills. Two examples of process goals are:

"Students will increase self esteem through goal setting." and "Girls will be encouraged to take part in higher level math and science courses."

Goals for school or organizational outcomes are related to improvement of the school but, unlike indicators and processes, are not directly connected to student learning. Therefore, it is even more difficult to tie them to the active use of skills and knowledge by students. Two examples of organizational outcomes are: "School employees will feel positive, empowered, and their accomplishments will be recognized." and "Parents and community members will become significantly involved in the achievement of vision, goals, and objectives."

Goals that are unclear don't stand on their own as explanations for students' use of knowledge and skills. Therefore, it is impossible to determine whether the school is attempting to improve something of value. Three examples of unclear goals that indicate target areas relating to complex behaviors follow.

"Children will say positive things about each other."

"Students will demonstrate rational and irrational thinking."

"Art - all areas and all techniques."

Goal Categorization by the Level of Learning Being Addressed

Each goal also was classified by the level of student learning it described. Three categorizations were used: integrated learning requiring active use of skills and knowledge, skill or concept acquisition requiring recall, and unclear. Examples of goals follow.

- Integrated learning:

  "Students will apply their knowledge of historical, economic, political, and geographic patterns to analyze five themes in American History: war, peace, expansion, depression, and globalization."

  "Students will master the use of tools to construct and repair objects they use in everyday life."

"Students will develop the ability to analyze problems in various situations and curricular areas through the use of concrete mathematical models."

- Skill or content acquisition:
  "Students will use "I" statements rather than engage in name calling."

  "The writer will use proper punctuation on written assignments and tests in English."

  "Students will locate and interpret information using developmentally appropriate resource materials."

- Unclear:
  "Students will understand the consequences of their actions."

  "The student's self-esteem will improve."

  "80% of the students will increase written communication skills 80% of the time by achieving at or above grade level norms on portfolio assignments."

  "Students will learn to teach themselves."

Goal Categorization for Measurability

Goals that are clearly measurable in terms of student learning outcomes describe the active use of knowledge and skills that can be observed and recorded, perceived and reported, or documented by artifacts. The following are examples of clearly measurable goals written by OA candidacy applicants:

"Students will use technological resources to improve oral and written communications in all academic areas."

"Students will demonstrate self-discipline to improve their social skills specifically when: a. interacting together , b. accepting responsibility, c. coping with criticism."

"Children will demonstrate the ability to work cooperatively in diverse groupings for various activities related to the school curriculum and its community."

If a goal cannot be authenticated by observable use of student knowledge or skills and the school supplied no data justifying its selection, it is considered immeasurable. The following four examples are of goals with measurement problems.

"Students will reflect higher morals on the moral response surveys."

"Students will become more responsible by increasing attendance and reducing late homework assignments."

"Students will follow rules: Students will become successful later in life."

"Citizenship proficiency scores will increase to 85%."

Goal Categorization for Equity

Within-school equity requires that all students be given the same expectations for improved learning. Therefore, equitable goals address improved learning outcomes for all of the school's students. The three goals that follow are equitable because they apply to every student in the school.

> "All students will demonstrate the ability to vary their writing style including vocabulary and sentence structure for different readers and different purposes."
>
> "Students will understand and demonstrate an acceptance of individual in social and academic situations."
>
> " Students will build the skills they need to become independent, life-long learners who are able use inquiry to acquire the knowledge they need."

Goals that specifically include or inadvertently exclude persons or groups of students have equity issues. If a target for a goal is not applicable to all of the school's students then it is not considered equitable. Following are examples for each type of coding for inequitable expectations.

- Academic achievement:
  "Ninth and tenth grade students will show an increase in the numbers of students passing classes."
- Socioeconomic status
  "Free lunch students will show an increase in achievement scores."
- Student behavior:
  "Student incidents of cheating will be reduced."

Appendix II

MATCH TO TEMPLATE (MTT) OF A SCHOOL'S SET OF GOALS

A "fit to template" is used to assess all OA candidate sets of goals for compliance with OA and to compare reviewers' independent judgments and the averaged peer reviews to the OA diagnostic template. An example of a school's OA improvement plan is used here. It would receive a rating of 70% for MTT, or 2.8 on conversion to a four-point scale. Table 1A provides an analysis of the plan by the goals included in it. It forms the basis of the score for the degree of MTT. Table 1B, which follows, provides a summary of the content analysis of the goals. This summary is applied to the diagnostics by assessing the extent to which the goals in the plan address the constructs for each diagnostic criteria.

SAMPLE SET OF OUTCOMES ACCREDITATION
SCHOOL IMPROVEMENT GOALS

I. All students will increase their respect for others and for property.

II. Students in speech class will demonstrate their ability to take a topic, choose and organize related ideas, and present their ideas clearly in standard English for the purpose of speaking to a group.

III. Students will demonstrate an improvement in test-taking skills.

IV. Students will use mathematical and scientific concepts in all curricular areas.

V. Students will develop the critical thinking skills necessary to develop solutions for problems in various mathematical settings.

Table 1A Content Analysis Of Individual Goals To Build A MTT

GOAL	Learning outcome	Complex behavior	Focus	Level of learning	Data-based	Equity
I.	yes	Caring for Self & Others	Citizenship	integrated use	yes	yes
II.	yes	Communicating	Speaking	acquisition	yes	no
III.	yes	Problem Solving	Study Skills	acquisition	yes	unsure
IV.	yes	Problem Solving	Problem Solving Thinking	integrated	yes	yes
V.	yes	Problem Solving	Problem Solving Skill	acquisition	yes	unsure

Table 1B Summary Analysis Of The School's Goals As An Entire Set

	Number	Percent
Total number of goals in plan	5	100%
Number of SLO goals	5	100%
Number of different complex behaviors	3	60%
Number of different curricular focuses	4	80%
Number of goals that integrate learning	2	40%
Number of goals that can be documented by data	5	100%
Number of goals that are clearly equitable	2	40%

From information in Table 1B, a MTT score for the entire set of goals is developed in the following manner:

1. SLOs: 100%
2. Equity: 40%

3. Data-based: 100% = (100% + 100%)/2 or (% SLOs + % measurable)/2

4. Challenge: 80% = (100% + 100% + 40%)/3 or (% SLOs + % measurable + % integrated learning)/3

5. Higher level skills: 73% = (100% + 80% + 40%)/3 or (% SLOs + % different complex behaviors + % integrated learning)/3

6. Staff involvement: 53% = (40% + 40% + 80%/3) or (% different curricular foci uses + % integrated learning + % equitable)/3

7. Coordination and integration: 47% = (60% + 40% + 40%)/3 or (% different complex behaviors + % integrated learning + % equitable)/3

Finally, the degree to which each construct is met is divided by the total number of constructs for the OA diagnostics (seven), resulting in a score for MTT; or the degree to which this school's plan matches the OA template. The resulting overall MTT of the sample set of school improvement goals used in the example is 0.70 or 70%. To convert this score to a four-point scale, like the holistic decision rating, it is multiplied by 4, equaling 2.8. The process for this particular set of goals can be condensed into:

$$MTT = ([100 + 40 + 100 + 80 + 73 + 53 + 47] / 7) \times 4$$

In summary, the MTT is derived by going through:

a content analysis of the goals in a school's OA plan (Table 1A); determining the number and percentage of goals in the plan that are SLOs, complex behaviors, curricular foci uses, integrated learning factors, suggest equitable expectations for students, and include data-based element (Table 1B); and determining the percentage to which the set of goals meets the OA template diagnostics converted to a four-point scale.